Topological Echoes of Primordial Physics in the Universe at Large Scales

Alex Cole
Gravitation Astroparticle Physics Amsterdam (GRAPPA)
Institute for Theoretical Physics Amsterdam
University of Amsterdam, the Netherlands
a.e.cole@uva.nl

Matteo Biagetti
IFPU Trieste, Italy
matteo.biagetti@sissa.it

Gary Shiu
Department of Physics
University of Wisconsin-Madison, USA
shiu@physics.wisc.edu

Abstract

We present a pipeline for characterizing and constraining initial conditions in cosmology via persistent homology. The cosmological observable of interest is the cosmic web of large scale structure, and the initial conditions in question are non-Gaussianities (NG) of primordial density perturbations. We compute persistence diagrams and derived statistics for simulations of dark matter halos with Gaussian and non-Gaussian initial conditions. For computational reasons and to make contact with experimental observations, our pipeline computes persistence in sub-boxes of full simulations and simulations are subsampled to uniform halo number. We use simulations with large NG ($f_{\rm NL}^{\rm loc} = 250$) as templates for identifying data with mild NG ($f_{\rm NL}^{\rm loc} = 10$), and running the pipeline on several cubic volumes of size $40\ ({\rm Gpc/h})^3$, we detect $f_{\rm NL}^{\rm loc} = 10$ at 97.5% confidence on $\sim 85\%$ of the volumes for our best single statistic. Throughout we benefit from the interpretability of topological features as input for statistical inference, which allows us to make contact with previous first-principles calculations and make new predictions.

1 Introduction

The last few decades have established a concordance model of cosmology, which relies on a handful of very well-constrained parameters [1, 2]. Despite this progress, mysteries remain. Perhaps the most pressing question is the mechanism that seeded primordial perturbations. These perturbations evolved under gravity to form the cosmic structure we observe today. By studying the statistics of cosmological observables, we aim to reverse-engineer the theory that generated primordial perturbations. Observations of the cosmic microwave background have so far provided the strongest constraints on the two-point function of primordial perturbations [3], which however only provides coarse-grained information about the free theory of the fields generating these perturbations, while not answering fundamental questions about the field content and interactions during this early phase. To probe the precise microphysics of this theory, we must move beyond the two-point function. One way to do this is to search for evidence of primordial non-Gaussianity (NG), the precise form of which allows one to discriminate between different theories of the very early universe [4]. For example, local NG [5] is described by the ansatz

$$\Phi_{\rm NG}(\mathbf{x}) = \Phi_{\rm G}(\mathbf{x}) + f_{\rm NL}^{\rm loc}\left(\Phi_{\rm G}^2(\mathbf{x}) - \left\langle \Phi_{\rm G}^2(\mathbf{x})\right\rangle\right)^2 \tag{1}$$

Topological Data Analysis and Beyond Workshop at the 34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada.

where $\Phi_{\rm NG}$ is the primordial gravitational potential, $\Phi_{\rm G}$ is a Gaussian random field, and $f_{\rm NL}^{\rm loc}$ controls the amount of NG. Single field models necessarily produce very small local NG as implied by consistency relations in the squeezed limit [6, 7], though large NG can be realized in other shapes [8]. Primordial NG has still not been detected in any form. Of the various forms of NG, local NG is the most tightly constrained, with $f_{\rm NL}^{\rm loc} = -0.9 \pm 10.2$ at 2σ [9].

There are several reasons to expect that topology is sensitive to NG. Namely, departures from Gaussianity are expected via an Edgeworth expansion to be most noticeable in the tail of a probability distribution. In terms of the primordial density perturbation, these tails are probed by local minima and local maxima. Under gravity, local maxima (overdense regions) collapse to form dark matter halos while local minima (underdense regions) expand as cosmic voids. These objects, along with filament loops, form a topological decomposition of large-scale structures (LSS) in the universe.

We therefore propose to robustly compute the topology of LSS and corresponding constrain cosmology using persistent homology [10, 11, 12]. In the context of LSS, persistent homology tracks the creation and destruction of clusters, filament loops, and voids across scales. While certain qualitative aspects of persistent homology have been computed before for LSS [13, 14, 15], the ability to perform statistics and generate constraints is underexplored. As the first work in this vein, we take into account several important observational effects, including varying halo density. Additionally, this is the first computation of the persistent topology of LSS with non-Gaussian initial conditions. In a cosmology context, the statistics derived from persistent homology have the advantage over other ML techniques of interpretability: we are able to reproduce previous results in certain limits and we make new predictions for unexplored observables, such as filament loops formed by dark matter halos.

This is a condensed version of the paper [16].

2 Method

Data set, subboxing, subsampling

We use the EOS DATASET[1], a suite of N-body simulations created to investigate the imprint of primordial non-Gaussianity in LSS at low redshift. The nomenclature for the cosmologies in the dataset is as follows: (G/NG)#(L/S), where G/NG distinguishes Gaussian from non-Gaussian initial conditions, # gives the value of the cosmological parameter σ_8 in the G case or the value of $f_{\rm NL}^{\rm loc} \neq 0$ in the NG case (NG simulations have $\sigma_8 = 0.85$, the same σ_8 value as in G85 simulations), and L/S determines whether a simulation was computed in a "large box" of side length of $2\ \mathrm{Gpc/h}$ or a "small box" of side length of $1\ \mathrm{Gpc/h}$ with different resolution. Each simulation in L has $\mathcal{O}(10^6)$ DM halos at redshift $z = 1$. When performing statistics, we will consider concatenated simulations with volume $40\ (\mathrm{Gpc/h})^3$. In the end we are considering similar volumes, redshift ranges, and DM halo mass ranges as those measured by upcoming galaxy surveys like Euclid. One difference to keep in mind is that galaxy surveys observe luminous galaxies hosted by DM halos, while restricting to DM-only simulations allows us to maintain control over the relevant physics.

The size of individual realizations in the EOS DATASET presents a challenge for persistent homology. To deal with it, we use filtrations built from subcomplexes of the Delaunay triangulation, as well as computing filtrations and their persistent homology in *sub-boxes* of a full simulation with side length $1\ \mathrm{Gpc/h}$. This is possible without introducing systematics because the topological features in our filtrations die at scales of $\mathcal{O}(10)\mathrm{Mpc/h} \ll 1\ \mathrm{Gpc/h}$.

An additional complication in the context of experimental observations of LSS is cosmic variance, which means that the number of halos in a fixed simulation volume can vary between simulations. This effect can compete with variations in halo density due to cosmology, and can have a large effect on persistence diagrams. To remove this systematic, we *subsample* each simulation to have the same number of halos. Without subsampling, two cosmologies that increase the total number of halos are much more difficult to distinguish via persistent homology, while subsampled simulations can be distinguished [16].

[1] Information on the full dataset is available at https://mbiagetti.gitlab.io/cosmos/nbody/eos/. The original paper investigated the imprint of primordial non-Gaussianity on the halo power spectrum [17].

Filtrations and statistics

Our primary filtration is a modification of the DTM-based filtration of [18]. In particular, we consider subcomplexes of the Delaunay triangulation organized by both scale and the extent to which a particular point is an outlier. We call this an αDTMℓ-filtration. The outlier property of a point is quantified by the Distance-to-Measure (DTM) function. Given a set of points X, a positive integer k, and a positive real number p, the empirical DTM function is given by [19]

$$\mathrm{DTM}(x) = \frac{1}{k}\left(\sum_{X_i \in N_k(x)} ||x - X_i||^p\right)^{1/p} \tag{2}$$

where $N_k(x)$ is the list of the k nearest neighbors in X to x. We will take $p = 2$ and $k = 15$. Now, given a point $x \in X$ and a positive real number q we consider the radius function

$$r_x(\nu) = (\nu^q - \mathrm{DTM}(x)^q)^{1/q} \quad \text{if } \nu \geq \mathrm{DTM}(x) \tag{3}$$

and $-\infty$ otherwise, and by definition a ball of radius $r = -\infty$ is the empty set. The role of $\mathrm{DTM}(x)$ is to delay the contribution of outliers to the filtration. Here q controls the mixing between $\mathrm{DTM}(x)$ and ν. We will take $q = 2$. For very large q, $r_x(\nu)$ is $-\infty$ until $\nu = \mathrm{DTM}(x)$, at which point it quickly jumps to ν. Beyond the vertices, we add an edge $[x_1 x_2]$ that is present in the Delaunay complex once the relevant balls overlap, i.e. $B_{r_{x_1}(\nu)} \cap B_{r_{x_2}(\nu)} \neq \emptyset$. Triangles and tetrahedra are then added when all necessary faces are present. In this construction, topological features will suffer less from the presence of outliers that in the α-filtration.

From the persistence diagrams corresponding to these filtrations, we derive several summary statistics for the purpose of inference. These include Persistence Images (PIs) [20] and the "topological curves" defined by

$$B_p(\nu) = \text{count}_p(\nu > \nu_{\mathrm{birth}}) \tag{4}$$
$$D_p(\nu) = \text{count}_p(\nu > \nu_{\mathrm{death}}) \tag{5}$$
$$P_p(\nu) = \text{count}_p(\nu > \nu_{\mathrm{persist}}) \tag{6}$$
$$b_p(\nu) = D_p(\nu) - B_p(\nu) \tag{7}$$

The curves B_p, D_p, P_p correspond up to normalization to empirical distribution functions of births, deaths, and persistences of p-cycles in a given diagram, while b_p are the familiar Betti numbers. Note that once a binning in ν is specified, these curves live in finite-dimensional vector spaces.

3 Results

We show the average PIs and their differences for our simulations in Figure 1. We examine the signature generated by $f_{\mathrm{NL}}^{\mathrm{loc}} = 250$, as the signal is large and results in clean patterns when averaged over our simulation volume. In the end we are interested in "detecting" $f_{\mathrm{NL}}^{\mathrm{loc}} = 10$ using these patterns. The effects of a nonzero $f_{\mathrm{NL}}^{\mathrm{loc}}$ on all the topological features share a common turnover feature: non-Gaussian simulations dominate the statistics for small and large birth scales while at intermediate scales, Gaussian simulations dominate. The PIs suggest that the topological curves $B_p(\nu)$, $D_p(\nu)$, and $b_p(\nu)$ should outperform $P_p(\nu)$ in detecting $f_{\mathrm{NL}}^{\mathrm{loc}}$.

Recall that 0-cycles in our filtrations are born at the DTM values of their respective halos. Examining PI_0, we see that $f_{\mathrm{NL}}^{\mathrm{loc}} > 0$ mainly causes a population of 0-cycles to shift to earlier births, i.e. smaller DTM values, while a subdominant population shifts to larger DTM values. This is consistent with the heuristic observation that positive $f_{\mathrm{NL}}^{\mathrm{loc}}$ affects structure as if there were more gravitational evolution in overdense regions (lowering the DTM values of corresponding halos as the neighborhood becomes more dense with halos) and less in underdense regions (increasing DTM values as fewer halos are formed there) [21, 22, 23, 24].

Among the topological curves we consider, D_p is correlated with the sizes of topological features. We show the effect of $f_{\mathrm{NL}}^{\mathrm{loc}} > 0$ on D_p in Fig. 2. Examining these more closely, we find that D_2 exhibits features consistent with previously derived results [25]. Ref. [25] performed a first-principles calculation of the effect of $f_{\mathrm{NL}}^{\mathrm{loc}}$ of the distribution of void sizes, finding that $f_{\mathrm{NL}}^{\mathrm{loc}}$ would increase the

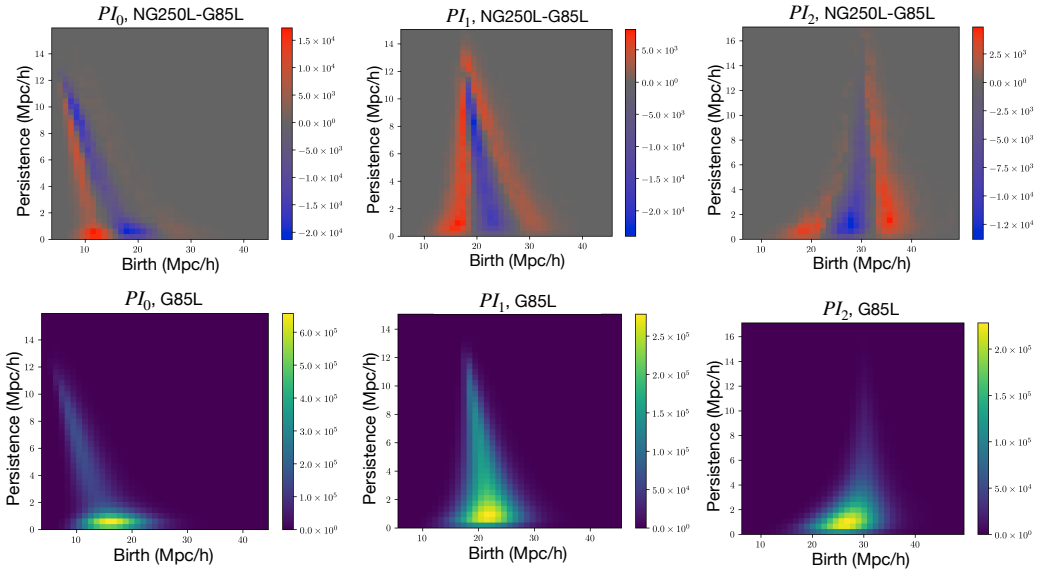

Figure 1: The effect of local non-Gaussianity on PIs for the αDTMℓ-filtration. In the top row the difference between the PIs averaged over NG250L and G85L is shown. The bottom row shows the PIs averaged over G85L.

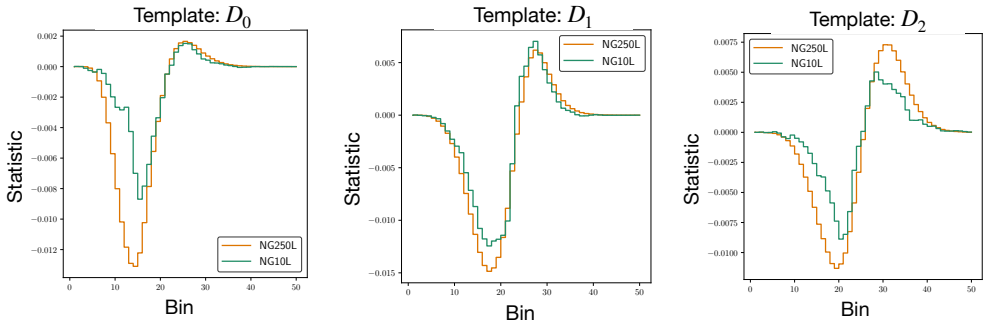

Figure 2: The effect of $f_{\rm NL}^{\rm loc} > 0$ on D_p. We show $D_p(f_{\rm NL}^{\rm loc} > 0) - D_p(f_{\rm NL}^{\rm loc} = 0)$ averaged over realizations with non-Gaussian simulations taken either from NG10L or NG250L and Gaussian simulations taken from G85L.

number of voids of small radius and decrease the number of voids of large radius. The manifests as the initial dip and subsequent rise of $D_2(f_{\rm NL}^{\rm loc} > 0) - D_2(f_{\rm NL}^{\rm loc} = 0)$ in Fig. 2. At large scales, we again measure a larger fraction of voids dying in the non-Gaussian case, a prediction not made in the approximation scheme of [25].

Intriguingly, a similar structure in D_1 suggests that similar physics is at play for filament loops. However, for these observables first-principles results are lacking, in part because filament loops break spherical symmetry. Our results motivate the development of analytical techniques like those in [25] for studying filament loops.

All in all, the interpretability of our statistics allows us to make contact with previously derived quantities, while motivating future study to derive other aspects of our results from first principles.

We can perform the following statistical test: for a statistic $\vec{S}$ (e.g. B_1, here written as a vector), we compute the *template* $\vec{T}_{\rm NG} \equiv \vec{S}_{\rm NG250S} - \vec{S}_{\rm G85S}$, where the subscript denotes averaging over all realizations. Given a simulation to test for NG, we compute the topological statistic $\vec{S}_{\rm test}$ and then

compute the distance

$$D_{\text{template}} = \frac{\vec{S}_{\text{test}} \cdot \vec{T}_{\text{NG}} - \vec{S}_{\text{G85L}} \cdot \vec{T}_{\text{NG}}}{\sigma} \tag{8}$$

where σ describes the variance of $\vec{S} \cdot \vec{T}_{\text{NG}}$ within G85L. Intuitively, if the test simulation has NG initial conditions, D_{template} will be positive via the construction of $\vec{T}_{\text{NG}}$. To generate quantitative predictions, we set a baseline by using test simulations drawn from G85L. This case models our uncertainty due to cosmic variance: even if some actual galaxy survey data had the exact same cosmological parameters as G85L, it would still be seeded by a different realization of a random field.

We run this pipeline with test simulations taken from G85L and for NG10L (being careful to use distinct simulation seeds). We use the distribution of D_{template} when the test data comes from G85L to set a 97.5% detection threshold for NG10L, see Fig. 3. Each data point in Fig. 3 reoresents a combination of simulations with total volume $40\ (\text{Gpc/h})^3$. We find that for the statistic B_1, NG at a level $f_{\text{NL}}^{\text{loc}} = 10$ is detected at 97.5% confidence for $\sim 85\%$ of these volumes.

The standard approach to detecting NG in LSS relies on the two- and three-point correlations of galaxies (or in our case DM halos). Recently, a rigorous study of these correlation functions for the EOS DATASET was not able to detect $f_{\text{NL}}^{\text{loc}} = 10$ at 2σ [26]. Our current topological methods are sensitive to higher-order correlations in the data, allowing them to probe more subtle patterns in the distribution of halos.

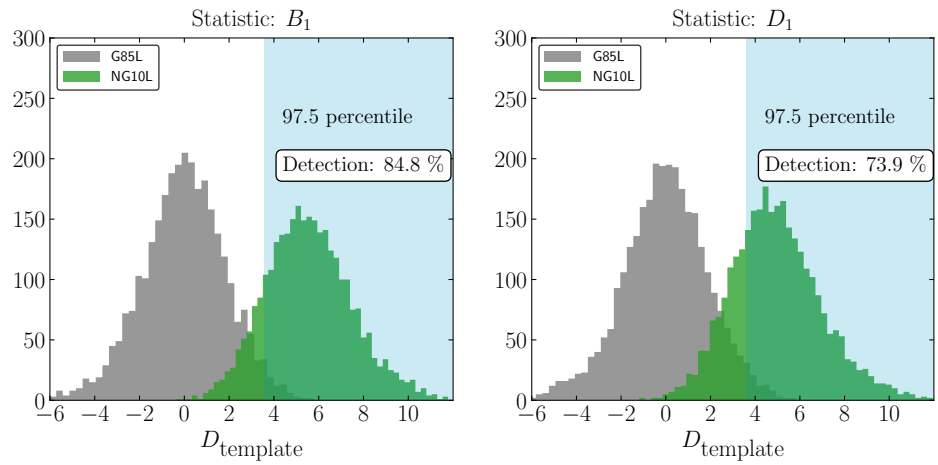

Figure 3: Distribution of D_{template}. Each data point represents a volume of $40\ (\text{Gpc/h})^3$. The fraction of NG10L data points lying beyond the 97.5% threshold of the G85L distribution is quoted as the detection rate of $f_{\text{NL}}^{\text{loc}} = 10$.

4 Discussion

We have introduced several practical considerations for persistent homology including subsampling and the use of sub-boxes to generate quantitative constraints on cosmology within a set of simulations. While this pipeline was presented in terms of detecting primordial non-Gaussianity, it is applicable to constraining any cosmological parameter. In future work we plan to study the effect of a broad range of cosmological parameters on these topological observables. We plan to eventually apply this pipeline to galaxy survey data, for which systematics like baryon feedback and redshift space distortions must be taken into account.

Our code, which proves a convenient API for loading halo catalogs and computing their persistent topology, summary statistics, and corresponding constraints, is provided publicly.[2]

[2]The code is available at https://gitlab.com/mbiagetti/persistent_homology_lss.

References

[1] N. Aghanim et al. Planck 2018 results. VI. Cosmological parameters. 7 2018.

[2] Shadab Alam et al. The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Cosmological Implications from two Decades of Spectroscopic Surveys at the Apache Point observatory. 7 2020.

[3] Y. Akrami et al. Planck 2018 results. X. Constraints on inflation. 2018.

[4] N. Bartolo, E. Komatsu, Sabino Matarrese, and A. Riotto. Non-Gaussianity from inflation: Theory and observations. *Phys. Rept.*, 402:103–266, 2004.

[5] Eiichiro Komatsu and David N. Spergel. Acoustic signatures in the primary microwave background bispectrum. *Phys. Rev.*, D63:063002, 2001.

[6] Juan Martin Maldacena. Non-Gaussian features of primordial fluctuations in single field inflationary models. *JHEP*, 05:013, 2003.

[7] Paolo Creminelli and Matias Zaldarriaga. Single field consistency relation for the 3-point function. *JCAP*, 0410:006, 2004.

[8] Xingang Chen, Min-xin Huang, Shamit Kachru, and Gary Shiu. Observational signatures and non-Gaussianities of general single field inflation. *JCAP*, 0701:002, 2007.

[9] Y. Akrami et al. Planck 2018 results. IX. Constraints on primordial non-Gaussianity. *Astron. Astrophys.*, 641:A9, 2020.

[10] Herbert Edelsbrunner and John Harer. *Computational topology: an introduction*. American Mathematical Soc., 2010.

[11] Afra J Zomorodian. *Topology for computing*, volume 16. Cambridge university press, 2005.

[12] Steve Y Oudot. *Persistence theory: from quiver representations to data analysis*, volume 209. American Mathematical Society Providence, 2015.

[13] Rien van de Weygaert et al. Probing Dark Energy with Alpha Shapes and Betti Numbers. 10 2011.

[14] Rien van de Weygaert et al. Alpha, Betti and the Megaparsec Universe: on the Topology of the Cosmic Web. *Trans. Comput. Sci.*, 14:60–101, 2011.

[15] Pratyush Pranav, Herbert Edelsbrunner, Rien van de Weygaert, Gert Vegter, Michael Kerber, Bernard J. T. Jones, and Mathijs Wintraecken. The Topology of the Cosmic Web in Terms of Persistent Betti Numbers. *Mon. Not. Roy. Astron. Soc.*, 465(4):4281–4310, 2017.

[16] Matteo Biagetti, Alex Cole, and Gary Shiu. The Persistence of Large Scale Structures I: Primordial non-Gaussianity. 9 2020.

[17] Matteo Biagetti, Titouan Lazeyras, Tobias Baldauf, Vincent Desjacques, and Fabian Schmidt. Verifying the consistency relation for the scale-dependent bias from local primordial non-Gaussianity. *Mon. Not. Roy. Astron. Soc.*, 468(3):3277–3288, 2017.

[18] Hirokazu Anai, Frédéric Chazal, Marc Glisse, Yuichi Ike, Hiroya Inakoshi, Raphaël Tinarrage, and Yuhei Umeda. DTM-based filtrations. *arXiv e-prints*, page arXiv:1811.04757, Nov 2018.

[19] Frédéric Chazal, David Cohen-Steiner, and Quentin Mérigot. Geometric inference for probability measures. *Foundations of Computational Mathematics*, 11(6):733–751, 2011.

[20] Henry Adams, Tegan Emerson, Michael Kirby, Rachel Neville, Chris Peterson, Patrick Shipman, Sofya Chepushtanova, Eric Hanson, Francis Motta, and Lori Ziegelmeier. Persistence images: A stable vector representation of persistent homology. *The Journal of Machine Learning Research*, 18(1):218–252, 2017.

[21] Neal Dalal, Olivier Doré, Dragan Huterer, and Alexander Shirokov. The imprints of primordial non-gaussianities on large-scale structure: scale dependent bias and abundance of virialized objects. *Phys. Rev.*, D77:123514, 2008.

[22] Xi Kang, P. Norberg, and J. Silk. Can large-scale structure probe CMB-constrained non-Gaussianity? *Mon. Not. Roy. Astron. Soc.*, 376:343–347, 2007.

[23] M. Grossi, E. Branchini, K. Dolag, S. Matarrese, and L. Moscardin. The mass density field in simulated non-Gaussian scenarios. *Mon. Not. Roy. Astron. Soc.*, 390:438–446, 2008.

[24] Annalisa Pillepich, Cristiano Porciani, and Oliver Hahn. Universal halo mass function and scale-dependent bias from N-body simulations with non-Gaussian initial conditions. *Mon. Not. Roy. Astron. Soc.*, 402:191–206, 2010.

[25] Marc Kamionkowski, Licia Verde, and Raul Jimenez. The Void Abundance with Non-Gaussian Primordial Perturbations. *JCAP*, 0901:010, 2009.

[26] Azadeh Moradinezhad Dizgah, Matteo Biagetti, Emiliano Sefusatti, Vincent Desjacques, and Jorge Noreña. Primordial Non-Gaussianity from Biased Tracers: Likelihood Analysis of Real-Space Power Spectrum and Bispectrum. 10 2020.